ASSIGNING STATISTICAL SIGNIFICANCE TO TUMOR CHANGES IN PATIENT MONITORING USING FDG PET

Perrine Tylski[1], Michelle Dusart[2], Bruno Vanderlinden[2], Irène Buvat[1]

[1]IMNC – U8165 CNRS – Paris 7 – Paris 11, Orsay, France
[2] Institut Jules Bordet, Bruxelles, Belgium

ABSTRACT

In PET-based patient monitoring, tumor changes can be assessed using standardized uptake values (SUV), tumor volume (V), or total lesion glycolysis (TLG). We studied the impact of the SUV, V and TLG estimation methods on the interpretation of tumor changes between 2 PET scans. We also propose a bootstrap approach to assign statistical significance to the observed tumor changes.

In 17 tumor changes, the SUV variations were the least dependent on the estimation method compared to the V or TLG changes. In 16/17 cases, SUV changes were significant. In 2 out of these 16 significant cases, at least one SUV index suggested non significant change.

Testing the significance of tumor feature changes might reduce errors in interpreting tumor changes.

Index Terms— Positron emission tomography, lung, tumors, biomedical monitoring, statistics

1. INTRODUCTION

FDG PET is increasingly used for therapy monitoring in oncology [1]. Yet, there is no consensus on the way tumor volume and metabolism should be measured from the PET images. The purpose of this study was to determine the impact of the measurement method on the data interpretation in the context of patient monitoring. We also propose a practical method to assign statistical significance to the observed tumor changes.

2. MATERIALS AND METHODS

2.1. Patient data

Two patients with lung cancer were scanned on a GE Discovery LS PET/CT system 5 and 6 times over the course of chemotherapy (from March 2006 to January 2007 and from December 2005 to February 2007). PET images were reconstructed using OSEM (2 iterations, 28 subsets, Gaussian post-smoothing of 5.45 mm full width at half maximum), corrected for attenuation using the CT and for scatter using a convolution subtraction method. Eight lesions in total were followed from one scan to the next, yielding 17 tumor changes between two consecutive scans available for analysis.

2.2. Segmentation methods

Four methods were used to segment the tumors on the PET data: manual delineation (Expert), two threshold-based methods (Tmax and Tcon) and an original fitting method (Fit), giving 4 independent volume estimates.

2.2.1. Expert delineation

The tumors were manually delineated by an expert using the PET images only, without considering the CT images or the PET/CT image fusion.

2.2.2. Maximum Intensity Threshold method

Tmax is a conventional threshold method in which all neighbor pixels with an intensity equal to or greater than a certain percentage of the maximum pixel intensity in the tumor are considered as belonging to the tumor [2]. The threshold t_{max} is defined in Eq. 1, where I_{max} is the maximum pixel intensity and α was taken equal to 0.4.

$$t_{\max} = \alpha \times I_{\max} \qquad \text{Eq.1}$$

2.2.3. Contrast-based Threshold method

Tcon is a threshold method taking into account the contrast between the tumor and the background intensity. The threshold t_{con} is defined by:

$$t_{con} = \beta_{opt} \times I_{70\%} + I_{background} \qquad \text{Eq.2}$$

$I_{70\%}$ is the mean intensity in a region including all pixels with an intensity equal or greater than 70% of the maximum pixel intensity in the tumor. $I_{70\%}$ is used instead of $I_{\max}$ as it is less influenced by noise. β_{opt} was determined using a phantom experiment. 6 hollow spheres with volumes from 0.5 to 16 mL were placed in a cylindrical phantom (Figure 1). The spheres and the phantom were filled with FDG to get sphere-to-background activity ratios of 10. β_{opt} was calculated as the average of the 6 β yielding a region with a volume equal to the volume of each sphere. The parameter β_{opt} was found to be equal to 0.336.

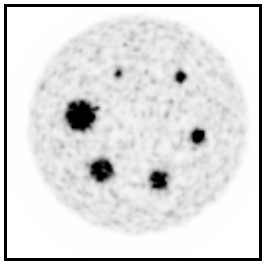

Figure 1. Cylindrical phantom with the 6 spheres

2.2.4. Fitting method

The fitting method, derived from Chen et al [3], used a 3D geometric model based on a "tumor" shape derived from activity thresholding. The spatial resolution in the reconstructed images was also assumed to be known. While Chen et al assumed a spherical shape for the tumor (which is a crude approximation for most tumors), in our approach, the tumor 3D shape was first determined using the Tcon method with a low threshold, defined as:

$$t_{low} = (\beta_{opt} - 0.05) \times I_{70\%} + I_{background} \qquad \text{Eq.3}$$

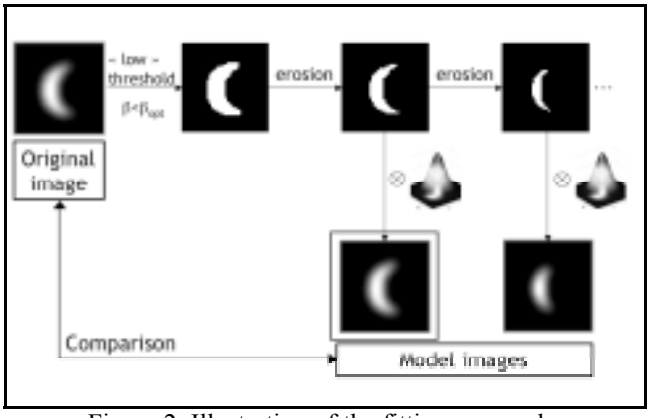

Figure 2. Illustration of the fitting approach.

The region segmented with the t_{low} threshold was assumed to be systematically larger than or equal to the actual tumor volume. In addition, it was assumed that the shape of this region was more similar to the shape of the tumor than a sphere.

Erosions of the resulting 3D shape were then performed to determine the volume that best fitted the observed image, using a 3D stationary convolution model based on the spatial resolution in the reconstructed image (Figure 2). A simplex algorithm determined the number of erosions, the activity values in the background and in the lesion minimizing a least square difference between the model and the image. Because the spatial resolution of the reconstructed images is embedded in the model, the estimated activity values were intrinsically corrected for partial volume effect. As the original voxel size in the image (3.9*3.9*4.25 mm^3) was too large to get precise volume estimates using successive erosions, pixel size was first reduced to 1*1*1 mm^3 with a cubic PCHIP interpolation [4,5] before applying the initial t_{low} threshold. The convolved model images were then resampled to the original pixel size to be compared to the original image. The fitting method was implemented in the MATLAB software (The MathWorks, Inc., Natick, MA).

2.3. SUV calculations

Six methods were used to estimate the SUV. SUVmax was defined as the maximum pixel value in the tumor. A mean SUV called SUV15mm was calculated in a VOI (Volume Of Interest) consisting of 3 circular 15-mm-diameter regions centered on the maximum intensity pixel in 3 consecutive slices. Two mean SUV were calculated in regions segmented according to the 2 previously described threshold segmentation methods (SUVmeanTmax, SUVmeanTcon). The fitting method gave an estimation of the SUV corrected for partial volume effect SUVmeanFit. We also calculated SUVmeanExpertRC, defined as the mean SUV in the VOI delineated by the expert corrected for partial volume effect using a recovery coefficient method [6,7].

2.4. TLG calculations

Total Lesion Glycolysis is defined as the product of SUV and volume and describes the global metabolism of the tumor [8].

For each of the 4 segmentation methods, the corresponding TLG was calculated, as the product of the volume by the corresponding SUV (TLGmeanTmax, TLGmeanTcon, TLGmeanFit, TLGmeanExpertRC).

2.5. Statistical analysis

2.5.1. Variability of the methods

To determine the impact of the estimation method on the measured change in feature estimates between two scans, we calculated, for each feature (SUV, V and TLG), the mean difference between the two extreme percent changes C_{min}(i) and C_{max}(i):

$$\Delta C = \frac{\sum_{i=1}^{N_{tc}} C_{\min}(i) - C_{\max}(i)}{N_{tc}} \qquad \text{Eq.4}$$

where N_{tc} is the number of tumor changes i considered, C_{min}(i) is the minimum percent change observed for tumor i for the considered feature (SUV, V, or TLG) and C_{max}(i) is the maximum percent change observed for tumor i for that feature.

The greater this mean difference, the larger the variability of the result as a function of the parameter estimation method.

2.5.2. Bootstrap analysis

As will be shown below, the measured change in tumor features between two scans highly depended on the method used to estimate the tumor feature.

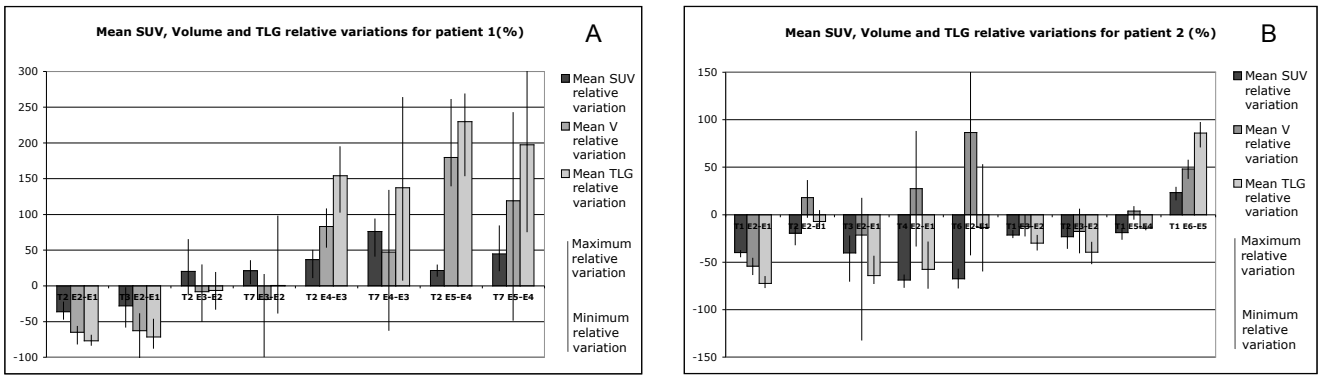

Figure 3. Mean estimated variation of SUV, Volume and TLG for all the estimates for patient 1 (A) and patient 2 (B). TX E(Y+1)-E(Y) indicates the change of tumor X between scan Y and scan Y+1.

We therefore propose a statistical approach to take advantages of the different feature estimates. The method tests the null hypothesis that there is no change occurring between the two scans (e.g. H0 = "percent SUV change is zero" when considering the SUV feature), using a non-parametric bootstrap approach adapted to small samples.
Based on a N-value sample of the feature of interest (N=6 for the SUV in our case), the method consists in drawing B bootstrap samples of N values each [9]. For each bootstrap sample, the bootstrap statistics

$$t(z^*) = \frac{\bar{z}^* - \bar{z}}{\bar{\sigma}^* / \sqrt{N}} \qquad \text{Eq.5}$$

is calculated, where $\bar{z}^*$ is the mean of the bootstrap sample, $\bar{\sigma}^*$ is standard deviation of the bootstrap sample, $\bar{z}$ is the mean of the original sample.
The resulting B $t(z^*)$ values are compared to t_{obs} given by:

$$t_{obs} = \frac{\bar{z} - \mu}{\sigma / \sqrt{N}} \qquad \text{Eq.6}$$

where $\bar{z}$ and σ are respectively the mean and the standard deviation of the original sample, μ is the value of tested mean. The achieved significance level (ASL) is defined as:

$$ASL = \frac{\#\{t(z^{*b}) \geq t_{obs}\}}{B} \qquad \text{Eq.7}$$

We finally compared the ASL to the chosen level of significance α. For our application, we chose B = 200, μ=0 and α=0.05.

3. RESULTS

3.1. Reproducibility of the parameters characterizing tumor changes

The mean ΔC (± 1 sd) between the extreme percent changes (Equation 4) was 29%±21% for SUV, 78%±84% for TLG and 92%±79% for V. This suggests that the most reproducible parameter to estimate tumor change was the SUV percent change. Percent changes in volumes largely varied depending on the segmentation method. As TLG uses the volume information, it was also affected by this large variability. Figure 3 shows the mean estimated variation of SUV, Volume and TLG for all the estimates, for the tumor and the successive scans of one patient. The error bar limits represent the extreme percent change (C_{min} and C_{max}).

3.3. Statistical analysis

The ASL obtained by the bootstrap analysis of mean test are shown in Table 1.
For the 17 tumor changes considered (Table 1), one case (Patient 1, tumor 2, change between scan 2 and scan 3) shows an ASL value greater than our chosen level of significance α.=0.05, suggesting no significant change of SUV between the two scans. For all other 16 cases, the changes in SUV were found to be significant.
By considering this result as the gold standard to establish whether the percent change in SUV was indeed significant or not, we could test how often wrong interpretations would have occurred if interpreting a single SUV index. Assuming that a percent change in SUV less than 10% would not be interpreted as a significant SUV change, we found that in 2 out of the 16 significant changes in SUV, considering a single SUV estimate would have led to a misinterpretation (no significant change): for tumor 3 of patient 1, the SUVmeanTcon percent change between scan 1 and scan 2 was less than 10%, and for tumor 7 of patient 1, the SUV15mm percent change between scan 2 and scan 3 was less than 10%, while the ASL value indicated significant changes in SUV. On the other hand, for the non significant change corresponding to an ASL>0.05 (tumor 2 of patient 1 for between scan 2 and scan 3), 5 out of 6 SUV change estimates were greater than 10% in absolute value, indicating a significant change.

Patient1	T2 E2-E1	T3 E2-E1	T2 E3-E2	T7 E3-E2	T2 E4-E3	T7 E4-E3	T2 E5-E4	T7 E5-E4	
ASL	0	0.01	*0.085*	0	0.01	0	0	0.025	
Patient2	**T1 E2-E1**	**T2 E2-E1**	**T3 E2-E1**	**T4 E2-E1**	**T6 E2-E1**	**T1 E3-E2**	**T2 E3-E2**	**T1 E5-E4**	**T1 E6-E5**
ASL	0	0.005	0.015	0	0	0	0.005	0.01	0

Table 1. ASL obtained with the bootstrap analysis for 8 tumor changes of patient 1 and for 9 tumor changes of patient 2. TX E(Y+1)-E(Y) indicates the change of tumor X between scan Y and scan Y+1.

4. DISCUSSION

In our study, we showed that SUV was the most reproducible parameter to estimate tumor changes during therapy monitoring. As previously described [10]; percent changes in SUV mostly showed the same trend for the different SUV estimates, but the amplitude of these changes was highly variable depending on the SUV estimate. This variability makes it difficult to establish criteria to objectively assess tumor response from FDG PET images.

Our results suggest that variations of metabolic tumor volumes as measured from successive PET scans highly depend on the method used to measure the tumor volume (Figure 3). Figure 3 shows that the tumor volume and SUV can even follow opposite variations: mean increase in SUV and mean decrease in tumor volumes in tumor 7 of patient 1 between scan 2 and 3 for instance. Part of these observations might be due to inaccurate tumor volume estimates, although three of the four methods we considered for the volume estimates are used in clinical practice. For instance, the same tumor had an estimated volume of 93.9 mL according to the expert delineation, 62.4 mL with the Tmax method and 75.8 mL with the Tcon method. Therefore, the metabolic tumor volume derived from the PET images should be interpreted with caution.

Recommendations have been published to guide the assessment of tumor response from PET images [11]. These recommendations are based on a threshold above which SUV percent changes are considered significant. The SUV is calculated in a fixed VOI delineated on the baseline scan. These recommendations are used in clinical studies [12]. Our study showed that a single index can lead to a erroneous interpretation of the tumor change, and propose a statistically-based approach to test whether the observed changes in SUV is significant or not, by accounting for the variability existing between different SUV estimates.

5. CONCLUSION

By taking advantage of the different methods that can be used to estimate tumor features (V, SUV or TLG) in FDG PET, we introduced a simple approach to test the significance of changes in tumor features between two successive scans. Our results suggest that this might reduce errors in interpreting tumor changes in the context of therapy monitoring.

6. REFERENCES

[1] H. Otsuka, N. Morita, K. Yamashita, et al., "FDG-PET/CT for cancer management", *J Med Invest*, Vol. 54, pp 195-199, 2007

[2] U. Nestle, S. Kremp, A. Schaefer-Schuler, et al., "Comparison of different methods for delineation of 18F-FDG PET-positive tissue for target volume definition in radiotherapy of patients with non-small cell lung cancer", *J Nucl Med*, Vol. 46, pp1342-1348, 2005

[3] C. Chen, R. Muzic, A. Nelson et al., "Simultaneous recovery of size and radioactivity concentration of small spheroids with PET Data", *J Nucl Med*, Vol. 40, pp 118-130, 1999

[4] F.N. Fritsch, F. N. and R. E. Carlson, "Monotone Piecewise Cubic Interpolation," *SIAM J. Numerical Analysis*, Vol. 17, pp 238-246, 1980

[5] D. Kahaner, , C. Moler and S. Nash, *Numerical Methods and Software*, Prentice Hall, 1988

[6] R.M. Kessler, J.R. Ellis Jr and M. Eden, "Analysis of emission tomographic scan data: limitations imposed by resolution and background." *J Comput Assist Tomogr*, Vol 8, pp 514-522, 1984

[7] M. Soret, SL. Bacharch and I. Buvat, "Partial-Volume Effect in PET Tumor Imaging" *J Nucl Med*, Vol. 48, pp 932-945, 2007

[8] S.M. Larson, Y. Erdi, T. Akhurst, et al., "Tumor treatment response based on visual and quantitative changes in global tumor glycolysis using FDG-PET imaging: the visual response score and change in total lesion glycolysis", *Clin Pos Imag,* Vol. 2, pp 159-171, 1999

[9] B. Efron, R.J. Tibshirani, *An introduction to the bootstrap*, Chapman & Hall, 1993

[10] N.C. Krak, O.S. Hoekstra, A.A. Lammertsma, "Measuring response to chemotherapy in locally advanced breast cancer: methodological considerations.", *Eur J Nucl Med Mol Imaging*; Vol. 31, pp 103-111,2004

[11] H. Young, R. Baum, U. Cremerius, et al. "Measurement of clinical and subclinical tumour response using [18F]-fluorodeoxyglucose and positron emission tomography: review and 1999 EORTC recommendations." European Organization for Research and Treatment of Cancer (EORTC) PET Study Group, *Eur J Cancer*. Vol. 35, pp 1773-1782, 1999

[12] S. Stroobants, J. Goeminne, M. Seegers 18FDG-Positron emission tomography for the early prediction of response in advanced soft tissue sarcoma treated with imatinib mesylate, *Eur J Cancer*, Vol. 39., pp 2012-2020